Chapter 6
Integration

Abstract The total amount of some quantity is an important and useful concept. We introduce the concept of the integral, the precise mathematical expression for total amount. The fundamental theorem of calculus tells us how the total amount is related to the rate at which that amount accumulates.

6.1 Examples of Integrals

We introduce the concept of the integral using three motivating examples of total amount: distance, mass, and area.

6.1a Determining Mileage from a Speedometer

At the beginning of Chap. 3, we investigated the relation between a car's odometer and speedometer. We showed that if the speedometer were broken, it would still be possible to determine the speed of the moving car from readings of the odometer and a clock. Now we investigate the inverse problem: how to determine the total mileage if we are given the speedometer readings at various times. We assume that we have at our disposal the *total record* of speedometer readings throughout the trip, i.e., that we know the speed of the car, f, as a function of time.

In the graph of f shown in Fig. 6.1, t is measured in hours, speed in miles per hour (mph). Our problem can be formulated as follows: *Given speed as function f of time, determine the distance covered during the time interval $[a,b]$.*

Let us denote the distance covered by

$$D(f,[a,b]).$$

P.D. Lax and M.S. Terrell, *Calculus With Applications*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-1-4614-7946-8_6,

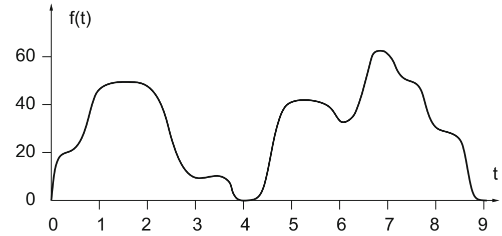

Fig. 6.1 Speed of a car

The notation emphasizes that D depends on f and $[a,b]$. How does D depend on $[a,b]$? Suppose we divide $[a,b]$ into two subintervals $[a,c]$ and $[c,b]$ that cover $[a,b]$ and do not overlap (Fig. 6.2).

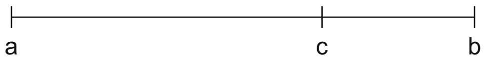

Fig. 6.2 The number c subdivides the interval

The distance covered during the total time interval $[a,b]$ is the sum of the distances covered during the intervals $[a,c]$ and $[c,b]$. This property is called the additivity property.

Additivity Property. For every c between a and b,

$$D(f,[a,b]) = D(f,[a,c]) + D(f,[c,b]). \tag{6.1}$$

How does distance D depend on the speed f? The distance covered by a car traveling with constant speed is

$$\text{distance} = (\text{speed})(\text{time}).$$

Suppose that between time a and b, the speed, f, is between m and M:

$$m \le f(t) \le M.$$

Two cars, one traveling with speed m the other with speed M, would cover the distances $m(b-a)$ and $M(b-a)$ during the time interval $[a,b]$. Our car travels a distance between these two. This property is called the lower and upper bound property.

Lower and Upper Bound Property. If $m \le f(t) \le M$ when $a \le t \le b$, then

$$m(b-a) \le D(f,[a,b]) \le M(b-a).$$

To make this example a bit more concrete, we use the data in Fig. 6.1 to find various lower and upper bounds for the distance traveled between $t = 2$ and $t = 7$.

From the graph, we see that the minimum speed during the interval was 0 mph and the maximum speed was 60 mph:

$$0 \le f(t) \le 60.$$

The time interval has length $7-2=5$ h. So we conclude that the distance traveled was between 0 and 300 miles,

$$0 = (0)(5) \le D(f,[2,7]) \le (60)(5) = 300.$$

This is not a very impressive estimate for the distance traveled. Let us see how to do better using additivity. We know that the time interval $[2,7]$ can be subdivided into two parts, $[2,5]$ and $[5,7]$, and that the total distance traveled is the sum of the distances traveled over these two shorter segments of the trip:

$$D(f,[2,7]) = D(f,[2,5]) + D(f,[5,7]).$$

The speed on $[2,5]$ is between 0 and 50 mph,

$$0 \le f(t) \le 50,$$

and the speed on $[5,7]$ is between 30 and 60 mph,

$$30 \le f(t) \le 60.$$

The length of $[2,5]$ is $5-2=3$, and the length of $[5,7]$ is $7-5=2$. The lower and upper bound property applied to each subinterval gives

$$(0)(3) \le D(f,[2,5]) \le (50)(3) \text{ and } (30)(2) \le D(f,[5,7]) \le (60)(2).$$

Adding these two inequalities together, we see that

$$60 \le D(f,[2,5]) + D(f,[5,7]) \le 270 \quad \text{miles.}$$

Now recalling that $D(f,[2,7]) = D(f,[2,5]) + D(f,[5,7])$, we get

$$60 \le D(f,[2,7]) \le 270 \quad \text{miles,}$$

a better estimate for the distance traveled.

6.1b Mass of a Rod

Picture a rod of variable density along the x-axis, as in Fig. 6.3. Denote the density at position x by $f(x)$ in grams per centimeter. Let $R(f,[a,b])$ denote the mass of the portion of the rod between points a and b of the x axis-measured in centimeters.

How does R depend on $[a,b]$? If we divide the rod into two smaller pieces lying on $[a,c]$ and $[c,b]$, then the mass of the whole rod is the sum of the mass of the pieces,

$$R(f,[a,b]) = R(f,[a,c]) + R(f,[c,b]).$$

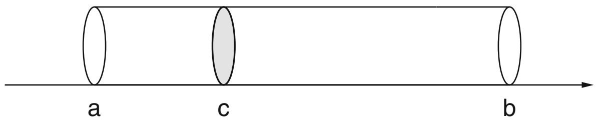

Fig. 6.3 A rod lying on a line between *a* and *b*

This property is the additivity property that we encountered in the distance example above.

How does the mass depend on f? If the density f is constant, then the mass is given by

$$\text{mass} \ = (\text{density})(\text{length}).$$

But our rod has variable density. If m and M are the minimum and maximum densities of the rod between a and b,

$$m \le f(x) \le M,$$

then the mass R of the $[a,b]$ portion of the rod is at least the minimum density times the length of the rod and is not more than the maximum density times the length of the rod:

$$m(b-a) \le R(f,[a,b]) \le M(b-a).$$

This is the lower and upper bound property that we encountered in the distance example.

Let us use the properties of additivity and lower and upper bounds to obtain various estimates for the mass of a particular rod. Suppose the rod lies along the x-axis between 1 and 5 cm and that its density at x is $f(x) = x$ grams per centimeter.

The greatest density occurs at $x = 5$ and is $f(5) = 5$, and the least density occurs at $x = 1$ and is $f(1) = 1$. The length of the rod is $5-1 = 4$ cm. So we conclude that

$$4 = (1)(4) \le R(f,[1,5]) \le (5)(4) = 20.$$

We can improve this estimate by subdividing the rod and using the properties of additivity and lower and upper bounds. This time, let us subdivide the rod into three shorter pieces as shown in Fig. 6.4.

By the additivity property, the mass of the rod between 1 and 5 is equal to the sum of the mass between 1 and 3 and the mass between 3 and 5. If we subdivide the

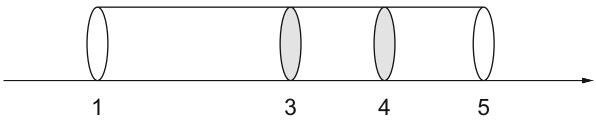

Fig. 6.4 A rod in three parts

interval $[3,5]$ at 4, then we see by additivity again that the mass between 3 and 5 is the sum of the mass between 3 and 4 and the mass between 4 and 5:

$$\begin{aligned} R(f,[1,5]) &= R(f,[1,3]) + R(f,[3,5]) \\ &= R(f,[1,3]) + R(f,[3,4]) + R(f,[4,5]). \end{aligned}$$

Next, we apply the lower and upper bound property to each segment to estimate their masses. The lower bounds for the density f on the three intervals are 1, 3, and 4, and the upper bounds are 3, 4, and 5, respectively. The lengths of the intervals are 2, 1, 1. By the lower and upper bound property, the masses of the three segments are

$$\begin{aligned} (1)(2) \le R(f,[1,3]) \le (3)(2), \\ (3)(1) \le R(f,[3,4]) \le (4)(1), \\ (4)(1) \le R(f,[4,5]) \le (5)(1). \end{aligned}$$

Now adding these three inequalities, we obtain

$$9 \le R(f,[1,3]) + R(f,[3,4]) + R(f,[4,5]) \le 15\,.$$

Recalling that mass has the additivity property

$$R(f,[1,5]) = R(f,[1,3]) + R(f,[3,4]) + R(f,[4,5]),$$

we obtain

$$9 \le R(f,[1,5]) \le 15,$$

a better estimate for the mass.

6.1c Area Below a Positive Graph

Let f be a function whose graph is shown in Fig. 6.5. We wish to calculate the area of the region contained between the graph of f, the x-axis, and the lines $x = a$ and $x = b$. Denote this area by

$$A(f,[a,b]).$$

How does A depend on $[a,b]$? For any c between a and b, subdivide $[a,b]$ into two subintervals $[a,c]$ and $[c,b]$, as on the left in Fig. 6.6. This subdivides the region

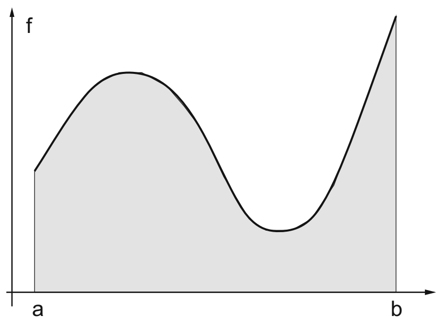

Fig. 6.5 The area of the shaded region is $A(f,[a,b])$

into two nonoverlapping regions. The area of the original region is the sum of the areas of the components. So A has the additive property,

$$A(f,[a,b]) = A(f,[a,c]) + A(f,[c,b]).$$

How does A depend on f? From the graph on the right in Fig. 6.6, we see that the values f takes on in $[a,b]$ lie between m and M:

$$m \le f(x) \le M \quad \text{for } x \text{ in } [a,b].$$

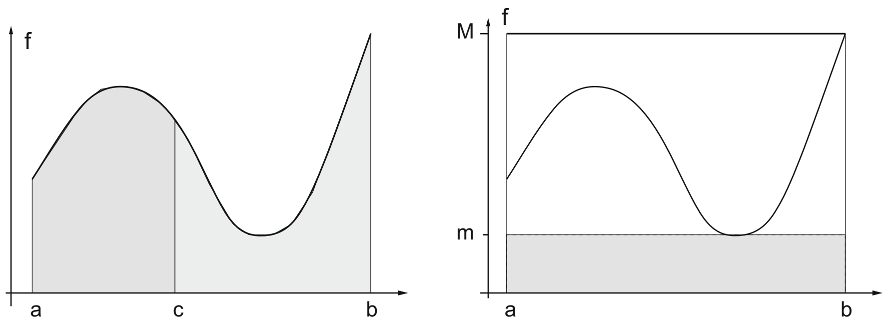

Fig. 6.6 *Left*: The interval subdivided. *Right*: Rectangles with heights m and M

Then, as Fig. 6.6 indicates, the region in question *contains* the rectangle with base $[a,b]$ and height m, and *is contained in* the rectangle with base $[a,b]$ and height M. Therefore, we conclude that

$$m(b-a) \le A(f,[a,b]) \le M(b-a). \tag{6.2}$$

That is, the lower and upper bound property holds.

Now just as we did in the examples of distance and mass, we look at a specific example. We estimate the area of the region bounded by the graph of $f(x) = x^2 + 1$, the x-axis, and the lines $x = -1$ and $x = 2$, as shown in Fig. 6.7. On $[-1,2]$, f is between 1 and 5, and so

$$3 = (1)(3) \le A(x^2+1,[-1,2]) \le (5)(3) = 15.$$

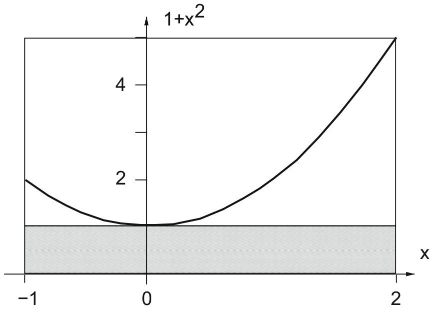

Fig. 6.7 The area below the graph of x^2+1 is between the areas of the smaller and larger rectangles. The height of the small rectangle is 1. The height of the large rectangle is 5

Since A is additive over intervals, we see that if we subdivide $[-1,2]$ into three intervals $[-1,-0.5]$, $[-0.5,1.5]$, and $[1.5,2]$, we get

$$A(x^2+1,[-1,2]) = A(x^2+1,[-1,-0.5]) + A(x^2+1,[-0.5,1.5]) + A(x^2+1,[1.5,2]).$$

Now on each of these subintervals, $f(x) = x^2+1$ takes on minimum and maximum values (see Fig. 6.8). So on $[-1,-0.5]$,

$$(1.25)(0.5) \le A(x^2+1,[-1,-0.5]) \le (2)(0.5).$$

On $[-.5,1.5]$,

$$(1)(2) \le A(x^2+1,[-0.5,1.5]) \le (3.25)(2);$$

and on $[1.5,2]$,

$$(3.25)(0.5) \le A(x^2+1,[1.5,2]) \le (5)(0.5).$$

Putting this all together, we get

$$\begin{aligned}&(1.25)(0.5) + (1)(2) + (3.25)(0.5)\\ &\le A(x^2+1,[-1,2]) \le (2)(0.5) + (3.25)(2) + (5)(0.5),\end{aligned}$$

or

$$4.25 \le A(x^2+1,[-1,2]) \le 10.$$

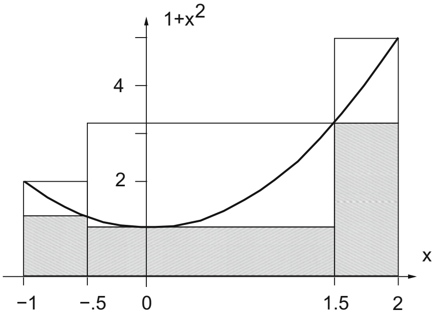

Fig. 6.8 The area below the graph of $x^2 + 1$ is between the sums of the areas of the smaller and larger rectangles. The heights of the small rectangles are 1.25, 1, 3.25. The heights of the large rectangles are 2, 3.25, 5

This is not a very accurate estimate for A, but it is better than our first estimate that A is between 3 and 15.

We can see from the graphs that if we were to continue to subdivide each of the intervals, the resulting estimates for A would become more accurate.

6.1d Negative Functions and Net Amount

So far, our function f has been positive, and in the case of density, that remains true. But signed distance D and area A also make sense for functions f which take on negative values.

The notion of positive and negative position of a car along a road can be defined in the same way as positive and negative numbers are defined on the number line: the starting point divides the road into two parts, one of which is arbitrarily labeled positive. Positions on the positive side are assigned a positive distance from the starting point, while positions on the negative side are assigned the negative of the distance from the starting point.

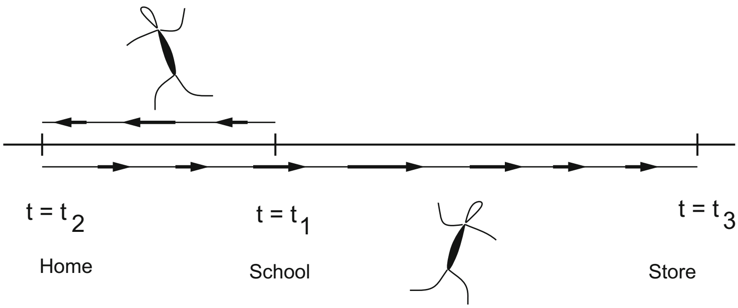

Fig. 6.9 A trip from school to home, then to the store

Velocity is defined as the derivative with respect to t of the position function as just defined. The change in position (or signed distance) traveled between two points in time is the ending position minus the starting position. On intervals over which velocity is always negative, the change in position (or signed distance) is negative. By arguments that are identical to those for positive velocity, we can see that signed distance D has both the additivity and the lower and upper bound properties. On intervals over which the velocity is at times positive and at times negative, we can see by additivity that D is the *net* distance: the sum of the signed distances traveled from the starting point. Fig. 6.9 shows an example.

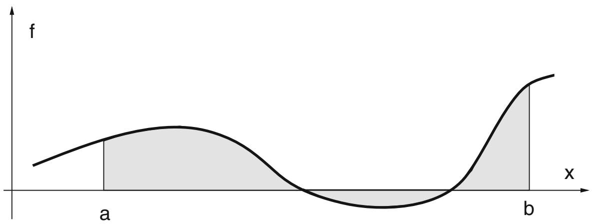

Fig. 6.10 A function f that has both positive and negative values

What should the interpretation of the area $A(f,[a,b])$ be when f takes on negative values, as pictured in Fig. 6.10? We propose to interpret the area above the x-axis as positive, and the area below the x-axis as negative, with the result that A will be defined as the algebraic sum of these positive and negative quantities. There is a reason for this interpretation: in many applications of area, the "underground" positions, i.e., those below the x-axis, have to be interpreted in a sense opposite to the portions aboveground. Only with this interpretation of positive and negative area is the lower and upper bound property (6.2) valid. This is particularly clear on intervals where f is negative.

We have shown that all three of the quantities distance D, rod mass R, and area A have the additive property with respect to $[a,b]$, and the lower and upper bound property with respect to f. We shall show in the next section that these two properties completely characterize D, R, and A. To put it sensationally, if you knew no more about D, R, and A than what you have learned so far, and if you were transported to a desert island, equipped only with pencil and paper, you could calculate the values of D, R, and A for any continuous function f on any interval $[a,b]$. The next section is devoted to explaining how.

Problems

6.1. Find a better estimate for the mass of the rod $R(x,[1,5])$ discussed in Sect. 6.1b, by

(a) Subdividing the rod into four subpieces of equal length,
(b) Subdividing the rod into eight subpieces of equal length.

6.2. Find better upper and lower estimates for the area $A(x^2+1,[-1,2])$ discussed in Sect. 6.1c, by subdividing $[-1,2]$ into six subintervals of equal length.

6.3. In this problem we explore signed area, or net area above the x-axis.

(a) Sketch the graph of $f(x)=x^2-1$ on the interval $[-3,2]$.
(b) Let $A(x^2-1,[a,b])$ be the signed area as described in Sect. 6.1d. Which of the following areas are clearly positive, clearly negative, or difficult to determine without some computation:

(i) $A(x^2-1,[-3,-2])$
(ii) $A(x^2-1,[-2,0])$
(iii) $A(x^2-1,[-1,0])$
(iv) $A(x^2-1,[0,2])$

(c) Find upper and lower estimates for $A(x^2-1,[-3,2])$ using five subintervals of equal length.

6.4. Let $f(t)=t^2-1$ be the velocity of an object at time t. Find upper and lower estimates for the change in position between time $t=-3$ and $t=2$ by subdividing $[-3,2]$ into five subintervals of equal length.

6.2 The Integral

We have seen that all three quantities distance $D(f,[a,b])$, rod mass $R(f,[a,b])$, and area $A(f,[a,b])$ are additive with respect to the given interval and have the lower and upper bound property with respect to f. In this section, we show that using only these two properties, we can calculate D, R, and A with as great an accuracy as desired.

In other words, if f and $[a,b]$ are the same in each of the three applications, then the numbers $D(f,[a,b])$, $R(f,[a,b])$, and $A(f,[a,b])$ have the same value, even though D, R, and A have entirely different physical and geometric interpretations. Anticipating this result, we call this number *the integral of f over* $[a,b]$ and denote it by

$$I(f,[a,b]).$$

The usual notation for the integral is

$$I(f,[a,b])=\int_a^b f(t)\,dt.$$

Example 6.1. The area $A(t,[0,b])$ is shown as the area of the large triangle in Fig. 6.11. In the integral notation it is written

$$\int_0^b t\,dt,$$

and since it represents the area of a triangle of base b and height b, the value is $\frac{1}{2}b^2$. According to the additive property, if $0 < a < b$, then

$$\int_0^b t\,\mathrm{d}t = \int_0^a t\,\mathrm{d}t + \int_a^b t\,\mathrm{d}t.$$

We find by subtracting that

$$\int_a^b t\,\mathrm{d}t = \frac{b^2 - a^2}{2}.$$

This is the area of the shaded trapezoid in the figure.

We use the notation $I(f, [a,b])$ when we are developing the concept of the integral to emphasize that it is an operation whose *inputs* are *a function and an interval* and whose *output* is a *number*. The basic properties of the integral are (a) additivity with respect to the interval of integration, and (b) the lower and upper bound property with respect to the function being integrated.

Next we give an example to show how we can compute an integral using only the two basic properties.

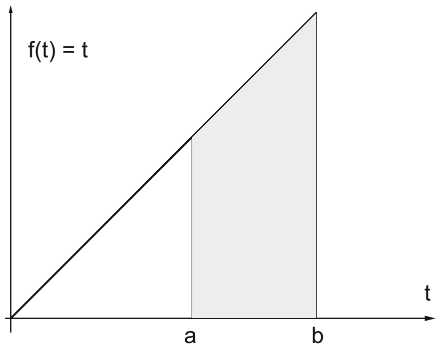

Fig. 6.11 The areas calculated in Example 6.1

The Integral of e^t on $[0,1]$: $\int_0^1 \mathrm{e}^t\,\mathrm{d}t$.

Let us look at what happens when we divide $[0,1]$ into three equal subintervals

$$0 < \frac{1}{3} < \frac{2}{3} < 1.$$

Since e^t is increasing, its lower bounds on these subintervals are

$$\mathrm{e}^0,\ \ \mathrm{e}^{1/3},\ \ \mathrm{e}^{2/3}.$$

Set $r = \mathrm{e}^{1/3}$. Then the lower bounds are $1,\ r,\ r^2$, and the upper bounds are $r,\ r^2,\ r^3$. The lower and upper bound property gives

$$(1)\left(\frac{1}{3}\right) \le \int_0^{1/3} \mathrm{e}^t\,\mathrm{d}t \le (r)\left(\frac{1}{3}\right),$$

$$(r)\left(\frac{1}{3}\right) \le \int_{1/3}^{2/3} \mathrm{e}^t\,\mathrm{d}t \le (r^2)\left(\frac{1}{3}\right),$$

$$(r^2)\left(\frac{1}{3}\right) \le \int_{2/3}^{1} \mathrm{e}^t\,\mathrm{d}t \le (r^3)\left(\frac{1}{3}\right).$$

Add these inequalities to get

$$\frac{1+r+r^2}{3} \le \int_0^{1/3} \mathrm{e}^t\,\mathrm{d}t + \int_{1/3}^{2/3} \mathrm{e}^t\,\mathrm{d}t + \int_{2/3}^{1} \mathrm{e}^t\,\mathrm{d}t \le \frac{r+r^2+r^3}{3}.$$

Then additivity gives

$$\frac{1+r+r^2}{3} \le \int_0^{1} \mathrm{e}^t\,\mathrm{d}t \le \frac{r+r^2+r^3}{3}.$$

Similarly, if the unit interval is divided into n equal parts, each of length $\frac{1}{n}$, we set $r = \mathrm{e}^{1/n}$. We get

$$\frac{1+r+\cdots+r^{n-1}}{n} \le \int_0^{1} \mathrm{e}^t\,\mathrm{d}t \le \frac{r+r^2+\cdots+r^n}{n}.$$

Both the left and right sides are $\frac{1}{n}$ times the partial sum of a geometric series (see Sect. 2.6a). We know that these sums can be rewritten as

$$\frac{1-r^n}{(1-r)n} \le \int_0^{1} \mathrm{e}^t\,\mathrm{d}t \le r\frac{1-r^n}{(1-r)n}. \tag{6.3}$$

Take $h = \frac{1}{n}$. Since $r = \mathrm{e}^{1/n}$, we have

$$\frac{1-r^n}{(1-r)n} = \frac{\mathrm{e}-1}{\frac{\mathrm{e}^h-1}{h}}.$$

As n tends to infinity, h tends to 0, and the limit $\lim_{h\to 0} \frac{\mathrm{e}^h-1}{h}$ is the derivative of e^x at $x = 0$, which equals 1. Therefore, in the limit as n tends to infinity, inequality (6.3) becomes $\mathrm{e}-1 \le \int_0^1 \mathrm{e}^t\,\mathrm{d}t \le \mathrm{e}-1$, and so

$$\int_0^{1} \mathrm{e}^t\,\mathrm{d}t = \mathrm{e}-1. \tag{6.4}$$

6.2a The Approximation of Integrals

We show now how to determine the integral of any continuous function over any closed interval, using only the following two basic properties:

$$\text{If } a < c < b \text{ then } \int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t \tag{6.5}$$

and

$$\text{If } m \le f(t) \le M \text{ then } m(b-a) \le \int_a^b f(t)\,\mathrm{d}t \le M(b-a). \tag{6.6}$$

By the additivity property in Eq. (6.5), we know that if

$$a < a_1 < b,$$

then

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^{a_1} f(t)\,\mathrm{d}t + \int_{a_1}^b f(t)\,\mathrm{d}t.$$

Similarly, divide $[a,b]$ into three intervals

$$a < a_1 < a_2 < b.$$

Then applying the additivity property twice, we get

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^{a_1} f(t)\,\mathrm{d}t + \int_{a_1}^b f(t)\,\mathrm{d}t \quad \text{and} \quad \int_{a_1}^b f(t)\,\mathrm{d}t = \int_{a_1}^{a_2} f(t)\,\mathrm{d}t + \int_{a_2}^b f(t)\,\mathrm{d}t.$$

Therefore,

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^{a_1} f(t)\,\mathrm{d}t + \int_{a_1}^{a_2} f(t)\,\mathrm{d}t + \int_{a_2}^b f(t)\,\mathrm{d}t.$$

Generally, if we divide $[a,b]$ into n intervals by

$$a < a_1 < a_2 < \cdots < a_{n-1} < b,$$

we find that repeated application of the additivity property gives

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^{a_1} f(t)\,\mathrm{d}t + \int_{a_1}^{a_2} f(t)\,\mathrm{d}t + \cdots + \int_{a_{n-1}}^b f(t)\,\mathrm{d}t. \tag{6.7}$$

Set $a_0 = a$ and $a_n = b$. Since f is continuous on $[a,b]$, we know by the extreme value theorem that f has a minimum m and maximum M on $[a,b]$, and a minimum m_i and a maximum M_i on each of the subintervals, with $m \le m_i$ and $M_i \le M$. We can estimate each of the integrals on the right-hand side of Eq. (6.7) by the lower and upper bound property:

$$m_i(a_i - a_{i-1}) \le \int_{a_{i-1}}^{a_i} f(t)\,\mathrm{d}t \le M_i(a_i - a_{i-1}).$$

By summing these estimates and using additivity, we get

$$\begin{aligned} &m_1(a_1-a_0)+m_2(a_2-a_1)+\cdots+m_n(a_n-a_{n-1}) \\ &\le \int_a^b f(t)\,\mathrm{d}t \le M_1(a_1-a_0)+M_2(a_2-a_1)+\cdots+M_n(a_n-a_{n-1}). \end{aligned} \tag{6.8}$$

Since each of the M_i is less than or equal to M, and $m \le m_i$, and since the sum of the lengths of the subintervals is the length $b-a$ of the entire interval, we see that inequality (6.8) is an improvement on the original estimate $m(b-a) \le \int_a^b f(t)\,\mathrm{d}t \le M(b-a)$. We saw examples of this kind of improvement in the distance, mass, and area examples.

Having a better estimate is good, but we would like to know that we can compute $\int_a^b f(t)\,\mathrm{d}t$ within any tolerance, no matter how small, through repeated uses of additivity and the lower and upper bound properties alone. We can achieve the tolerance we desire by making the difference between the left-hand side of Eq. (6.8), called the *lower sum*, and the right-hand side of Eq. (6.8), called the *upper sum*, as small as desired. For example, if the difference between them were less than $\frac{1}{1000}$, then we would know $\int_a^b f(t)\,\mathrm{d}t$ within that tolerance.

We recall now that every continuous function f on a closed interval is uniformly continuous. That is, given any tolerance $\varepsilon > 0$, there is a precision $\delta > 0$ such that if two points c and d in $[a,b]$ are closer than δ, then $f(c)$ and $f(d)$ differ by less than ε. So if we break up the interval $[a,b]$ into pieces $[a_{i-1},a_i]$ with length less than δ, the minimum m_i and the maximum M_i of f on $[a_{i-1},a_i]$ will differ by less than ε. The left and right sides of Eq. (6.8) differ by

$$(M_1-m_1)(a_1-a_0)+(M_2-m_2)(a_2-a_1)+\cdots+(M_n-m_n)(a_n-a_{n-1}),$$

which is less than

$$\varepsilon(a_1-a_0)+\varepsilon(a_2-a_1)+\cdots+\varepsilon(a_n-a_{n-1}) = \varepsilon(a_n-a_0) = \varepsilon(b-a).$$

This shows that for a sufficiently fine subdivision of the interval $[a,b]$, the upper sum and the lower sum in Eq. (6.8) differ by less than $\varepsilon(b-a)$, so by Eq. (6.8), the upper and lower sums differ from $\int_a^b f(t)\,\mathrm{d}t$ by less than $\varepsilon(b-a)$. Since $b-a$ is fixed, we can choose ε so that $\varepsilon(b-a)$ is as small as we like.

So far, our descriptions of how to determine the integral of a continuous function on $[a,b]$ have relied heavily on finding the maximum and minimum on each subinterval. In general, it is not easy to find the absolute maximum and minimum values of a continuous function on a closed interval, even though we know they exist. We now give estimates of the integral of a function that are much easier to evaluate than the upper and lower bounds in Eq. (6.8), and that also approximate $\int_a^b f(t)\,\mathrm{d}t$ within any tolerance.

Definition 6.1. Choose any point t_i in the interval $[a_{i-1}, a_i]$, $i = 1, 2, \dots, n$, and form the sum

$$I_{\text{approx}}(f, [a,b]) = f(t_1)(a_1 - a_0) + f(t_2)(a_2 - a_1) + \cdots + f(t_n)(a_n - a_{n-1}). \tag{6.9}$$

The sum I_{approx} is called an *approximate integral* of f on $[a,b]$ or a Riemann sum of f on $[a,b]$.

Example 6.2. To find an approximate integral of $f(t) = \sqrt{t}$ on the interval $[1,2]$ using the subdivision

$$1 < 1.3 < 1.5 < 2,$$

we need to choose a number from each subinterval. Let us choose them in such a way that their square roots are easy to calculate:

$$t_1 = 1.21 = (1.1)^2, \quad t_2 = 1.44 = (1.2)^2, \quad t_3 = 1.69 = (1.3)^2.$$

Then

$$\begin{aligned} I_{\text{approx}}(\sqrt{t}, [1,2]) &= \sqrt{t_1}(1.3 - 1) + \sqrt{t_2}(1.5 - 1.3) + \sqrt{t_3}(2 - 1.5) \\ &= (1.1)(0.3) + (1.2)(0.2) + (1.3)(0.5) = 1.22. \end{aligned}$$

Approximate integrals are easy to compute, but how close are they to $\int_a^b f(t)\,\mathrm{d}t$? The value $f(t_i)$ lies between the minimum m_i and the maximum M_i of f on $[a_{i-1}, a_i]$. Therefore, $I_{\text{approx}}(f, [a,b])$ lies in the interval between

$$m_1(a_1 - a_0) + m_2(a_2 - a_1) + \cdots + m_n(a_n - a_{n-1})$$

and

$$M_1(a_1 - a_0) + M_2(a_2 - a_1) + \cdots + M_n(a_n - a_{n-1}).$$

That is, no matter how the t_i are chosen in each interval, each approximate integral lies within the same interval that contains $\int_a^b f(t)\,\mathrm{d}t$. We saw that for continuous functions we can subdivide $[a,b]$ into subintervals such that the difference between the lower sum and the upper sum is less than $\varepsilon(b-a)$. Therefore, the exact and an approximate integral differ by an amount not greater than $\varepsilon(b-a)$. We state this result as the approximation theorem.

Theorem 6.1. Approximation theorem for the integral. *Suppose that for a continuous function f on $[a,b]$, $|f(c)-f(d)|<\varepsilon$ whenever $|c-d|<\delta$. Take a subdivision*

$$a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b, \tag{6.10}$$

where each length $(a_i - a_{i-1})$ is less than δ. Then every approximate integral

$$f(t_1)(a_1-a_0)+\cdots+f(t_n)(a_n-a_{n-1})$$

differs from the exact integral $\int_a^b f(t)\,\mathrm{d}t$ by less than $\varepsilon(b-a)$.

Sometimes, the length $(a_i - a_{i-1})$ of the ith interval in (6.10) is denoted by $\mathrm{d}t_i$, so that the approximating sums are written as

$$I_{\text{approx}} = f(t_1)(a_1-a_0)+\cdots+f(t_n)(a_n-a_{n-1}) = f(t_1)\mathrm{d}t_1+\cdots+f(t_n)\mathrm{d}t_n.$$

One term of this sum is illustrated in Fig. 6.12. If we use the sum symbol, this sum may be abbreviated as

$$I_{\text{approx}}(f,[a,b]) = \sum_{i=1}^{n} f(t_i)\,\mathrm{d}t_i. \tag{6.11}$$

We use the classical notation

$$\int_a^b f(t)\,\mathrm{d}t$$

for the integral because of its resemblance to this formula.

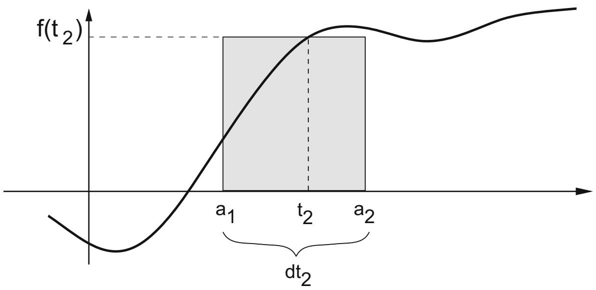

Fig. 6.12 Elements of an approximate integral

6.2b Existence of the Integral

We began our discussion of the integral with physical and geometric examples such as distance, mass, and area, in which it was reasonable to assume that there was a number, called the integral, that we were trying to estimate. For example, we found it reasonable to believe there is a number that can be assigned to the area of a planar region bounded by a nice boundary. But how do we know that a single number called area can be assigned to such a region in the first place?

In this section, we prove that for a continuous function on a closed interval, the approximate integrals converge to a limit as we refine the subdivision. We show that this limit does not depend on the particular sequence of subdivisions used. This limit is called the definite integral and is written $\int_a^b f(t)\,\mathrm{d}t$.

Given a continuous function f on $[a,b]$ and given any tolerance ε, we can choose δ such that if two points s and t in $[a,b]$ differ by less than δ, then $f(s)$ and $f(t)$ differ by less than ε. Subdivide the interval $[a,b]$ as

$$a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b.$$

Then choose any point t_i in the ith subinterval, $a_{i-1} \le t_i \le a_i$, and form the approximate integral $I = \sum_{i=1}^{n} f(t_i)(a_i - a_{i-1})$. Then form another approximate integral $I' = \sum_{j=1}^{m} f(t'_j)(a'_j - a'_{j-1})$ using another subdivision

$$a = a'_0 < a'_1 < \cdots < a'_m = b$$

and points t'_j from each subinterval. Next, we show that if all the lengths of the subintervals are small enough,

$$a_i - a_{i-1} < \frac{1}{2}\delta \qquad \text{and} \qquad a'_j - a'_{j-1} < \frac{1}{2}\delta,$$

then the two approximate integrals differ from each other by less than $\varepsilon(b-a)$. That is,

$$|I - I'| < \varepsilon(b-a).$$

To see this, we form the common subdivision consisting of all the intersections of the a intervals and a' intervals that have positive length. We denote by s_{ij} the length of the intersection of $[a_{i-1}, a_i]$ with $[a'_{j-1}, a'_j]$. An example is indicated in Fig. 6.13. Many of the s_{ij} are zero, because most of the subintervals do not overlap.

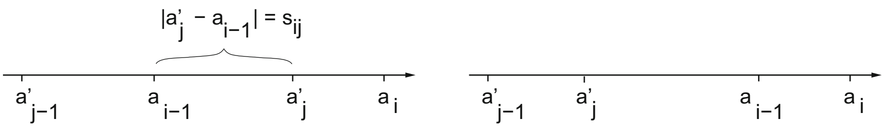

Fig. 6.13 *Left*: The subintervals intersect, with length s_{ij}. *Right*: They do not overlap, and $s_{ij} = 0$

We break up the sums I and I' as follows. The length $(a_i - a_{i-1})$ is the sum over j of the s_{ij} for which $[a_{i-1}, a_i]$ and $[a'_{j-1}, a'_j]$ intersect. Since the other s_{ij} are 0, we can sum over all i and j:

$$I = \sum_i f(t_i)(a_i - a_{i-1}) = \sum_{i,j} f(t_i)s_{ij}.$$

Similarly, the length $(a_j - a_{j-1})$ is the sum over i of the s_{ij} for which $[a_{i-1}, a_i]$ and $[a'_{j-1}, a'_j]$ intersect. So

$$I' = \sum_j f(t'_j)(a_j - a_{j-1}) = \sum_{i,j} f(t'_i)s_{ij}.$$

Therefore, $I - I' = \sum_{i,j}(f(t_i) - f(t'_j))s_{ij}$. The only terms of interest are those for which s_{ij} is nonzero, that is, for which $[a_{i-1}, a_i]$ and $[a'_{j-1}, a'_j]$ overlap. But the lengths of $[a_{i-1}, a_i]$ and $[a'_{j-1}, a'_j]$ have been assumed less than $\frac{1}{2}\delta$. Hence the points t_i and t'_j in each nonzero term differ by no more than δ. It follows that $f(t_i)$ and $f(t'_j)$ differ by less than ε. Therefore, by the triangle inequality,

$$|I - I'| \le \sum_{i,j} \varepsilon s_{ij} = \varepsilon(b - a). \tag{6.12}$$

This result enables us to make the following definition.

Definition 6.2. The integral of a continuous function on a closed interval. Take any sequence of subdivisions of $[a, b]$ with the following property: the length of the largest subinterval in the kth subdivision tends to zero as k tends to infinity. (For instance, we could take the kth subdivision to be the subdivision into k equal parts.) Denote by I_k any approximate integral using the kth subdivision.

Since f is a continuous function, given any tolerance $\varepsilon > 0$, there is a precision $\delta > 0$ such that the values of f differ by less than ε over any interval of length δ. Choose N so large that for $k > N$, each subinterval of the kth subdivision has length less than $\frac{1}{2}\delta$. It follows from (6.12) that for k and l greater than N, I_k and I_l differ by less than $\varepsilon(b - a)$. This proves the convergence of the sequence I_k.

The limit does not depend on our choice of the sequence of subdivisions. For given two such sequences, we can merge them into a single sequence, and the associated approximate integrals form a convergent sequence. This proves that the two sequences that were merged have the same limit.

This common limit is defined to be the integral $\int_a^b f(t)\,\mathrm{d}t$.

Other Integrable Functions. If f is not continuous on $[a,b]$, we may be able to define a continuous function g on $[a,b]$ by redefining f at finitely many points. If so, we say that f is integrable on $[a,b]$ and set $\int_a^b f(t)\,\mathrm{d}t = \int_a^b g(t)\,\mathrm{d}t$.

Example 6.3. To compute $\int_1^4 \frac{t^2-4}{t-2}\,\mathrm{d}t$, we notice that $\frac{t^2-4}{t-2}$ is not continuous on all of $[1,4]$ but is equal to $t+2$ for $t \neq 2$. So by redefining $\frac{t^2-4}{t-2}$ to be 4 at $t=2$, we obtain

$$\int_1^4 \frac{t^2-4}{t-2}\,\mathrm{d}t = \int_1^4 (t+2)\,\mathrm{d}t.$$

Example 6.4. To compute $\int_0^1 \frac{\sin t}{t}\,\mathrm{d}t$, we notice that $\frac{\sin t}{t}$ is not continuous at 0, because it is not defined at 0. We know that $\lim_{t\to 0} \frac{\sin t}{t} = 1$, so we define $g(0) = 1$ and $g(t) = \frac{\sin t}{t}$ for $t \neq 0$. Then g is continuous on every closed interval $[a,b]$. In particular, $\int_0^1 \frac{\sin t}{t}\,\mathrm{d}t = \int_0^1 g(t)\,\mathrm{d}t$ is a number. We do not have an easy way to calculate this number, but using ten equal subintervals and the right-hand endpoints, you can see that it is approximately

$$\int_0^1 \frac{\sin t}{t}\,\mathrm{d}t \approx \sum_{n=1}^{10} \frac{\sin\left(\frac{n}{10}\right)}{\frac{n}{10}} \left(\frac{1}{10}\right) \approx 0.94.$$

Also, if f is not continuous on $[a,b]$ but is integrable on $[a,c]$ and $[c,b]$, then we say that f is integrable on $[a,b]$ and set $\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t$.

Example 6.5. Denote by $[x]$ the greatest integer that is less than or equal to x. See Fig. 6.14. Then $[x]$ is integrable on $[0,1]$, $[1,2]$, and $[2,3]$, and

$$\int_0^3 [x]\,\mathrm{d}x = \int_0^1 0\,\mathrm{d}x + \int_1^2 1\,\mathrm{d}x + \int_2^3 2\,\mathrm{d}x = 0 + 1(2-1) + 2(3-2) = 3.$$

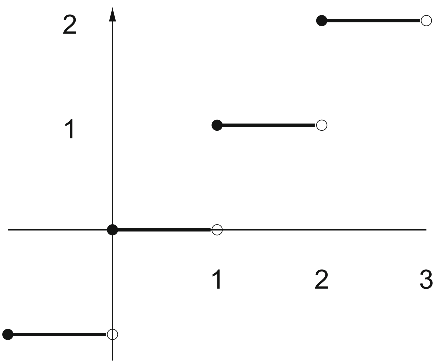

Fig. 6.14 Graph of the greatest integer function, in Example 6.5

The Properties of Integrals Revisited. We verify that the lower and upper bound property, as well as additivity property, are satisfied by Definition 6.2.

To check additivity, suppose $a < c < b$. Approximate $\int_a^b f(t)\,\mathrm{d}t$ using a subdivision of $[a,b]$ in which one point of the subdivision is c. This is permissible, because we have seen that any sequence of subdivisions is allowed, as long as the lengths of the subintervals tend to 0, and this can certainly be done while keeping c as one of the division points. Then for each such $I_{\text{approx}}(f,[a,b])$, we may separate the terms into two groups corresponding to the subintervals to the left of c and those to the right of c, and thus express $I_{\text{approx}}(f,[a,b])$ as a sum

$$I_{\text{approx}}(f,[a,b]) = I_{\text{approx}}(f,[a,c]) + I_{\text{approx}}(f,[c,b]).$$

As the lengths of the largest subintervals tend to zero, the approximate integrals tend to

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t.$$

This verifies the additivity property (6.5). For the lower and upper bound property, suppose $m \le f(t) \le M$ on $[a,b]$. Then for every approximate integral, we have

$$\sum_i m(a_i - a_{i-1}) \le \sum_i f(t_i)(a_i - a_{i-1}) \le \sum_i M(a_i - a_{i-1}),$$

that is, $m(b-a) \le I_{\text{approx}}(f,[a,b]) \le M(b-a)$. In the limit, it follows that

$$m(b-a) \le \int_a^b f(t)\,\mathrm{d}t \le M(b-a),$$

which is the lower and upper bound property (6.6).

6.2c Further Properties of the Integral

We present some important properties of the integral.

Theorem 6.2. The mean value theorem for integrals. *If f is a continuous function on $[a,b]$, then there is a number c in $[a,b]$ for which*

$$\int_a^b f(t)\,\mathrm{d}t = f(c)(b-a).$$

The number $f(c)$ is called the mean or average value of f on $[a,b]$.

Proof. By the extreme value theorem, f has a minimum value m and a maximum M on $[a,b]$. Then the lower and upper bound property gives

$$m \le \frac{1}{b-a}\int_a^b f(t)\,\mathrm{d}t \le M.$$

Since a continuous function takes on all values between its minimum and maximum, there is a number c in $[a,b]$ for which

$$f(c) = \frac{1}{b-a}\int_a^b f(t)\,\mathrm{d}t.$$

□

Example 6.6. The mean of $f(t) = t$ on $[a,b]$ is

$$\frac{1}{b-a}\int_a^b t\,\mathrm{d}t = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2} = f(c), \qquad \text{where} \quad c = \frac{a+b}{2}.$$

In Problem 6.9, we encourage you to explore how the mean value of f on $[a,b]$ is related to the ordinary average taken over n numbers, and to the concept of a weighted average.

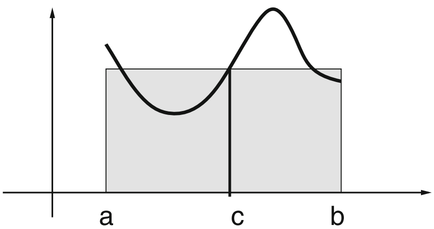

Fig. 6.15 The mean value of a positive function, illustrated in Example 6.7

Example 6.7. Take f positive. We interpret $\int_a^b f(t)\,\mathrm{d}t$ as the area under the graph of f in Fig. 6.15. Theorem 6.2 asserts that there is at least one number c between a and b such that the shaded rectangle shown has the same area as the region under the graph of f.

We make the following definition, which is occasionally useful in simplifying expressions.

Definition 6.3. When $a > b$ we define $\int_a^b f(t)\,\mathrm{d}t = -\int_b^a f(t)\,\mathrm{d}t$, and when $a = b$ we define $\int_a^b f(t)\,\mathrm{d}t = 0$.

Note that the additivity property holds for all numbers a, b, c within an interval on which f is continuous.

Example 6.8.

$$\int_1^3 f(t)\,\mathrm{d}t = \int_1^5 f(t)\,\mathrm{d}t + \int_5^3 f(t)\,\mathrm{d}t,$$

because this is merely a rearrangement of the earlier property

$$\int_1^3 f(t)\,\mathrm{d}t + \int_3^5 f(t)\,\mathrm{d}t = \int_1^5 f(t)\,\mathrm{d}t.$$

Even and Odd Functions. The integral of an odd function f, $f(-x) = -f(x)$, on an interval that is symmetric about 0 is zero. Consider approximate integrals for $\int_{-a}^0 f(x)\,\mathrm{d}x$ and $\int_0^a f(x)\,\mathrm{d}x$ for an odd function as in Fig. 6.16. We see that

$$\int_{-a}^0 f(x)\,\mathrm{d}x = -\int_0^a f(x)\,\mathrm{d}x.$$

Therefore, $\int_{-a}^a f(x)\,\mathrm{d}x = \int_{-a}^0 f(x)\,\mathrm{d}x + \int_0^a f(x)\,\mathrm{d}x = 0$ for odd f.

For an even function f, $f(-x) = f(x)$, as in the figure, we see that

$$\int_{-a}^0 f(x)\,\mathrm{d}x = \int_0^a f(x)\,\mathrm{d}x.$$

Therefore, $\int_{-a}^a f(x)\,\mathrm{d}x = 2\int_0^a f(x)\,\mathrm{d}x$ for even f.

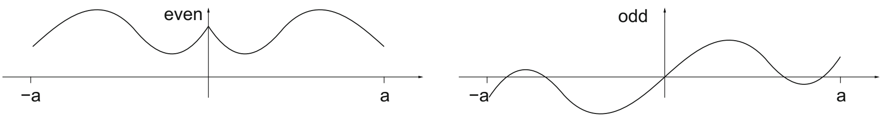

Fig. 6.16 Graphs of even and odd functions

Theorem 6.3. Linearity of the integral. *For any numbers a, b, c_1, c_2 and continuous functions f_1 and f_2, we have*

$$\int_a^b c_1 f_1(t) + c_2 f_2(t)\,\mathrm{d}t = c_1 \int_a^b f_1(t)\,\mathrm{d}t + c_2 \int_a^b f_2(t)\,\mathrm{d}t.$$

Proof. The approximate integrals satisfy

$$I_{\text{approx}}(c_1 f_1 + c_2 f_2, [a,b]) = c_1 I_{\text{approx}}(f_1, [a,b]) + c_2 I_{\text{approx}}(f_2, [a,b])$$

if we use the same subdivision and the same points t_i in each of the three sums. The limit of these relations then gives Theorem 6.3. □

Theorem 6.4. Positivity of the integral. *If f is a continuous function with $f(t) \geq 0$ on $[a,b]$, then $\int_a^b f(t)\,\mathrm{d}t \geq 0$.*

Proof. Each approximate integral consists of nonnegative terms, so the limit must be nonnegative. □

Example 6.9. If $f_1(t) \leq f_2(t)$ on $[a,b]$, then $\int_a^b f_1(t)\,\mathrm{d}t \leq \int_a^b f_2(t)\,\mathrm{d}t$. We see this by taking $f = f_2 - f_1$ in Theorem 6.4 and using the linearity of the integral, Theorem 6.3.

Problems

6.5. Calculate the approximate integral for the given function, subdivision of the interval, and choice of evaluation points t_i. For each problem, make a sketch of the graph of the function corresponding to the approximate integral.

(a) $f(t) = t^2 + t$ on $[1,3]$, using $1 < 1.5 < 2 < 3$ and $t_1 = 1.2$, $t_2 = 2$, $t_3 = 2.5$.
(b) $f(t) = \sin t$ on $[0,\pi]$, using $0 < \frac{\pi}{4} < \frac{\pi}{2} < \frac{3\pi}{4} < \pi$ and take the t_i to be the left endpoints of the subintervals.

6.6. Use an area interpretation of the integral to compute the integrals for the function f whose graph is shown in Fig. 6.17.

(a) $\int_0^1 f(t)\,\mathrm{d}t$
(b) $\int_1^4 f(t)\,\mathrm{d}t$
(c) $\int_0^4 f(t)\,\mathrm{d}t$
(d) $\int_1^6 f(t)\,\mathrm{d}t$

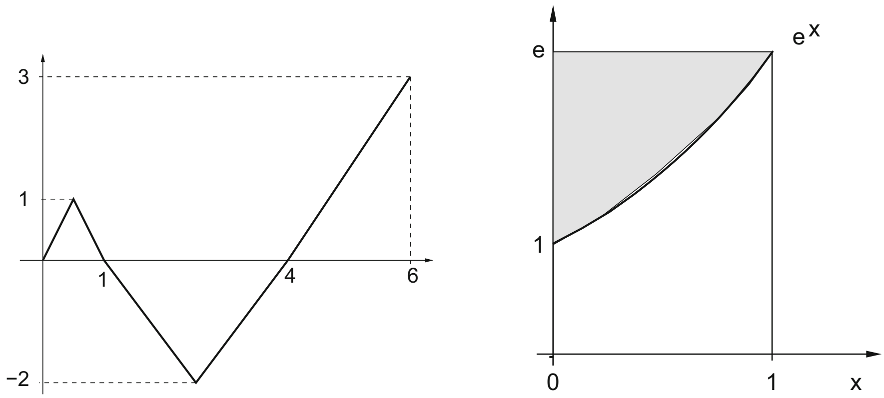

Fig. 6.17 *Left*: The graph of f in Problem 6.6. *Right*: The graph of e^x for Problem 6.7

6.7. Refer to Fig. 6.17.

(a) Use the result that $\int_0^1 \mathrm{e}^t\,\mathrm{d}t = \mathrm{e} - 1$ to find the area of the shaded region.
(b) Use a geometric argument to compute $\int_1^{\mathrm{e}} \log t\,\mathrm{d}t$.

6.8. Use an area interpretation and properties of integrals to evaluate the following integrals.

(a) $\int_{-\pi}^{\pi} \sin(x^3)\,\mathrm{d}x$
(b) $\int_0^2 \sqrt{4-x^2}\,\mathrm{d}x$
(c) $\int_{-10}^{10} \left(\mathrm{e}^{x^3} - \mathrm{e}^{-x^3}\right)\mathrm{d}x$

6.9. Let $T(x)$ be the temperature at point x on a rod, $0 \le x \le 15$.

(a) Use the measurements at six equally spaced points shown in Fig. 6.18. Make two estimates of the average temperature, first with an approximate integral using left endpoint values of T, and then using right endpoint values.
(b) Use the measurements taken at unequally spaced points in Fig. 6.18. Write two expressions to estimate the average temperature, first with an approximate integral using left endpoint values of T, and then using right endpoint values.

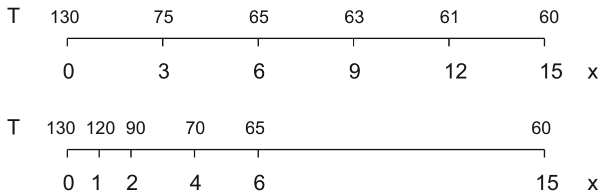

Fig. 6.18 Two sets of temperature measurements along the same rod in Problem 6.9

6.10. Express the limit

$$\lim_{n\to\infty} \frac{1+4+9+\cdots+(n-1)^2}{n^3}$$

as an integral of some function over some interval, and find its value.

6.11. Let k be some positive number. Consider the interval obtained from $[a,b]$ by stretching in the ratio $1:k$, i.e., $[ka,kb]$. Let f be any continuous function on $[a,b]$. Denote by f_k the function defined on $[ka,kb]$ obtained from f by stretching:

$$f_k(t) = f\left(\frac{t}{k}\right).$$

Using approximate integrals, prove that

$$\int_{ka}^{kb} f_k(t)\,\mathrm{d}t = k\int_a^b f(t)\,\mathrm{d}t$$

and make a sketch to illustrate this result. In Sect. 6.3, we shall prove this relation using the fundamental theorem of calculus.

Further properties of the integral are explored in the next problems. These properties can be derived here using approximate integrals. Later, in Sect. 6.3, we ask you to prove them using the fundamental theorem of calculus.

6.12. Let f be any continuous function on $[a,b]$. Denote by f_r the function obtained when f is shifted to the right by r. That is, f_r is defined on $[a+r,b+r]$ according to the rule

$$f_r(t) = f(t-r).$$

See Fig. 6.19. Prove that

$$\int_a^b f(x)\,\mathrm{d}x = \int_{a+r}^{b+r} f_r(x)\,\mathrm{d}x.$$

This property of the integral is called *translation invariance*.

Hint: Show that approximating sums are translation-invariant.

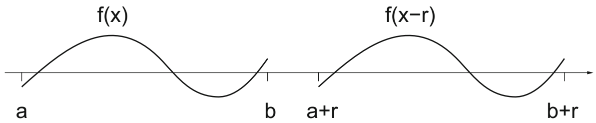

Fig. 6.19 Translation

6.13. For an interval $[a,b]$, the *reflected* interval is defined as $[-b,-a]$. If f is some continuous function on $[a,b]$, its *reflection*, denoted by f_-, is defined on $[-b,-a]$ as follows:

$$f_-(t) = f(-t).$$

The graph of f_- is obtained from the graph of f by reflection across the vertical axis; see Fig. 6.20. Prove that

$$\int_{-b}^{-a} f_-(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}t.$$

This property of the integral is called *invariance under reflection*.

Hint: Show that approximating sums are invariant under reflection.

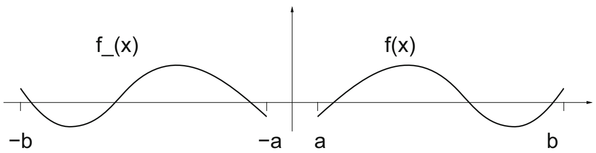

Fig. 6.20 Reflection

6.14. In this problem, you will evaluate $\int_0^1 \sqrt{t}\,\mathrm{d}t = \frac{2}{3}$.

(a) Using $0 < \frac{1}{4} < \frac{1}{2} < 1$, verify that the upper and lower sums are

$$I_{\text{upper}} = \sqrt{\frac{1}{4}}\frac{1}{4} + \sqrt{\frac{1}{2}}\frac{1}{4} + \sqrt{1}\frac{1}{2}$$

and

$$I_{\text{lower}} = \sqrt{0}\frac{1}{4} + \sqrt{\frac{1}{4}}\frac{1}{4} + \sqrt{\frac{1}{2}}\frac{1}{2}.$$

(b) Let $0 < r < 1$, and use the subdivision $0 < r^3 < r^2 < r < 1$. Verify that the upper sum is

$$I_{\text{upper}} = \sqrt{r^3}r^3 + \sqrt{r^2}(r^2 - r^3) + \sqrt{r}(r - r^2) + \sqrt{1}(1 - r),$$

and find an expression for the lower sum.

(c) Write the upper sum for the subdivision $0 < r^n < r^{n-1} < \cdots < r^2 < r < 1$, recognize a geometric series in it, and check that

$$I_{\text{upper}} \to \frac{1-r}{1-r^{3/2}}$$

as n tends to infinity.

(d) Show that $\frac{1-r}{1-r^{3/2}}$ tends to $\frac{2}{3}$ as r tends to 1.

6.3 The Fundamental Theorem of Calculus

Earlier, we posed the following problem: determine the change of position, or *net distance*, D, of a moving vehicle during a time interval $[a,b]$ from knowledge of the *velocity* f of the vehicle at each instant of $[a,b]$. The answer we found in Sect. 6.1d was that D is the integral

$$D = I(f,[a,b])$$

of the velocity as a function of time over the interval $[a,b]$. This formula expresses the net distance covered during the whole trip. A similar formula holds, of course, for the net distance $D(t)$ up to time t. This formula is

$$D(t) = I(f,[a,t]),$$

where $[a,t]$ is the interval between the starting time and the time t.

In Sects. 3.1 and 6.1d, we discussed the *converse problem*: if we know the net distance $D(t)$ of a moving vehicle from its starting point to its position at time t, for all values of t, how can we determine its velocity as a function of time

t? The answer we found was that *velocity is the derivative* of D as a function of time:

$$f(t) = D'(t)$$

We also posed the problem early in this chapter of finding the mass of a rod R between points a and b from knowledge of its linear density f. We found that the mass of the rod is

$$R = I(f,[a,b]),$$

the integral of the density f over the interval $[a,b]$. A similar formula holds for the mass of the part of the rod up to the point x. This is

$$R(x) = I(f,[a,x]).$$

Again in Sect. 3.1 we discussed the converse problem: if we know the mass of the rod from one end to any point x, how do we find the linear density of the rod at x? The answer we found was that the linear density f at x is the derivative of the mass

$$f(x) = R'(x).$$

We can summarize these observations in the following words:

If a function F is defined to be the integral of f from a to x,
then the derivative of F is f.

Omitting all qualifying phrases, we can express the preceding statement as an epigram.

Differentiation and integration are inverses of each other.

The argument presented in favor of this proposition was based on physical intuition. We proceed to give a purely mathematical proof.

Theorem 6.5. The fundamental theorem of calculus

(a) Let f be any continuous function on $[a,b]$. Then f is the derivative of some differentiable function. In fact, for x in $[a,b]$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^x f(t)\,\mathrm{d}t\right) = f(x). \tag{6.13}$$

(b) Let F be any function with a continuous derivative on $[a,b]$. Then

$$F(b) - F(a) = \int_a^b F'(t)\,\mathrm{d}t. \tag{6.14}$$

Proof. We first prove statement (a). Define a function

$$G(x) = \int_a^x f(t)\,dt.$$

Form the difference quotient

$$\frac{G(x+h) - G(x)}{h}.$$

We have to show that as h tends to zero, this quotient tends to $f(x)$. By definition of G,

$$G(x+h) = \int_a^{x+h} f(t)\,dt, \qquad \text{and} \qquad G(x) = \int_a^x f(t)\,dt.$$

By the additivity property of the integral,

$$\int_a^{x+h} f(t)\,dt = \int_a^x f(t)\,dt + \int_x^{x+h} f(t)\,dt.$$

This can be written as

$$G(x+h) = G(x) + \int_x^{x+h} f(t)\,dt.$$

The difference quotient is

$$\frac{G(x+h) - G(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,dt.$$

By the mean value theorem for integrals, Theorem 6.2, there is a number c between x and $x+h$ such that

$$\frac{1}{h}\int_x^{x+h} f(t)\,dt = f(c).$$

That is, as we see in Fig. 6.21, the area of a strip, divided by its width, is equal to the height of the strip at some point. Since f is continuous, $f(c)$ tends to $f(x)$ as h tends to 0. This proves that the difference quotient tends to $f(x)$. Therefore, the derivative of G is f. This concludes the proof of part (a).

We turn now to the proof of part (b). Since F' is continuous on $[a,b]$, we can define a function F_a by

$$F_a(x) = \int_a^x F'(t)\,dt, \quad a \le x \le b. \tag{6.15}$$

As we have shown in the proof of part (a), the derivative of F_a is F'. Therefore, the difference $F - F_a$ has derivative zero for every x in $[a,b]$. By Corollary 4.1 of the mean value theorem for derivatives, a function whose derivative is zero at every point of an interval is constant on that interval. Therefore,

$$F(x) - F_a(x) = \text{constant}$$

for every x. We evaluate the constant as follows. Let $x = a$. By the definition of F_a, $F_a(a) = 0$. Then $F(a) - F_a(a) = F(a) = $ constant, and we get

$$F(x) = F_a(x) + F(a) \quad \text{on} \quad [a,b].$$

Setting $x = b$, it follows that

$$F(b) - F(a) = F_a(b) = \int_a^b F'(t)\,\mathrm{d}t. \tag{6.16}$$

This completes a proof of the fundamental theorem of calculus. □

Here is another proof of part (b) of the fundamental theorem of calculus that uses the mean value theorem for derivatives more directly. Let

$$a = a_0 < a_1 < a_2 < \cdots < a_n = b$$

be any subdivision of the interval $[a,b]$. According to the mean value theorem for derivatives, in each subinterval $[a_{i-1}, a_i]$, there is a point t_i such that

$$F'(t_i) = \frac{F(a_i) - F(a_{i-1})}{a_i - a_{i-1}}.$$

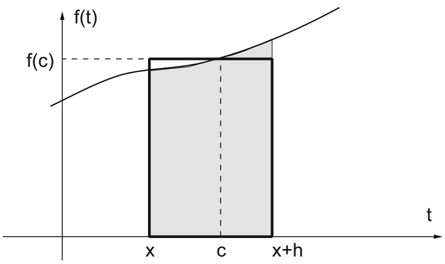

Fig. 6.21 The equation $\frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t = f(c)$ illustrated

Therefore,

$$F'(t_i)(a_i - a_{i-1}) = F(a_i) - F(a_{i-1}).$$

Add these up for all i from 1 to n. We get

$$\sum_{i=1}^{n} F'(t_i)(a_i - a_{i-1}) = F(a_1) - F(a_0) + F(a_2) - F(a_1) + \cdots + F(a_n) - F(a_{n-1})$$

$$= F(b) - F(a).$$

The sum on the left is an approximation to the integral $\int_a^b F'(t)\,\mathrm{d}t$. We have shown in Sect. 6.2b that the set of all approximations tends to the integral as the subdivision is refined. Our formula shows that no matter how fine the subdivision, *these particular* approximations are *exactly equal* to $F(b) - F(a)$. Therefore, the limit is $F(b) - F(a)$.

The fundamental theorem of calculus deserves its honorific name; it has at least two important uses. First and foremost, it is the *fundamental existence theorem* of analysis; it guarantees the existence of a function with a given derivative. Its second use lies in furnishing an exact method for evaluating the integral of any function we recognize to be the derivative of a known function.

We have shown how to deduce the fundamental theorem of calculus from the mean value theorem of differential calculus and the mean value theorem for integrals. The three can be related in one unifying statement. If F' is continuous on $[a,b]$, then for some c,

$$F'(c) = \frac{1}{b-a}\int_a^b F'(t)\,\mathrm{d}t = \frac{F(b)-F(a)}{b-a}.$$

We can express this relationship in words: the average of the instantaneous rates of change of F throughout an interval is equal to the average rate of change in F over the interval.

Notation. We sometimes denote $F(b)-F(a)$ by

$$F(b)-F(a) = [F(x)]_a^b = F(x)\Big|_a^b.$$

A function F whose derivative is f is called an *antiderivative* of f. One way to evaluate a definite integral on an interval is to find and evaluate an antiderivative F. We use the notation $\int f(x)\,\mathrm{d}x$ to denote an antiderivative of f. Often, $\int f(x)\,\mathrm{d}x = F(x)+C$ is used to denote all possible antiderivatives of f. Antiderivatives expressed in this way are called *indefinite* integrals. The constant C is called the constant of integration, and it can be assigned any value.

Evaluation of Some Integrals. Next, we illustrate how to use the fundamental theorem of calculus to evaluate the integral of a function that we recognize as a derivative.

Example 6.10. Since $\big(-\cos t\big)' = \sin t$, the fundamental theorem gives

$$\int_a^b \sin t\,\mathrm{d}t = -\cos t\Big|_a^b = -\cos b + \cos a.$$

Example 6.11. We know that $(\mathrm{e}^t)' = \mathrm{e}^t$. By the fundamental theorem, then,

$$\int_0^1 \mathrm{e}^t\,\mathrm{d}t = \mathrm{e}^1 - \mathrm{e}^0 = \mathrm{e} - 1,$$

in agreement with our computation of this integral in Sect. 6.2.

Example 6.12. Let $f(t) = t^c$, where c is any real number except -1. Then f is the derivative of $F(t) = \dfrac{t^{c+1}}{c+1}$. By the fundamental theorem,

$$\int_a^b t^c \, dt = F(b) - F(a) = \frac{b^{c+1}}{c+1} - \frac{a^{c+1}}{c+1}.$$

In particular,

$$\int_a^b t \, dt = \frac{t^2}{2}\bigg|_a^b = \frac{b^2 - a^2}{2},$$

as we found in Sect. 6.2.

By now, you must have noticed that the key to evaluating integrals by the fundamental theorem lies in an ability to notice that the function f presented for integration is the derivative of another handy function F. How does one acquire the uncanny ability to find antiderivatives? It comes with the experience of differentiating many functions; in addition, the search for F can be systematized with the aid of a few basic techniques. These will be presented in the next chapter.

The Special Case $c = -1$. We showed in Chap. 3 that for every number c, positive or negative except -1, the function t^c, $t > 0$, is the derivative of $\dfrac{t^{c+1}}{c+1}$. In contrast, the function t^{-1} is the derivative of $\log t$. This seems very strange. The functions t^c change continuously with c as c passes through the value -1. Why is there such a drastic discontinuity at $c = -1$ of the antiderivative of the function?

We show now that the discontinuity is only apparent, not real. Let

$$F_c(t) = \int_1^t x^c \, dx.$$

Using the fundamental theorem, we get

$$F_c(t) = \frac{t^{c+1} - 1}{c+1} \quad \text{for } c \neq -1, \qquad \text{and } F_{-1}(t) = \log t.$$

We shall show that as c tends to -1, F_c tends to F_{-1}. This is illustrated in Fig. 6.22. To prove this, we set $c = -1 + y$. Then

$$F_{y-1}(t) = \frac{t^y - 1}{y}.$$

Define the function g as $g(y) = t^y$, where t is some positive number. Note that $g(0)=1$. Using this function, we rewrite the previous relation as

$$F_{y-1}(t) = \frac{g(y) - g(0)}{y}.$$

The limit of this expression as y tends to zero is the derivative of g at $y = 0$. To evaluate that derivative, we write g in exponential form: $g(y) = e^{y \log t}$. Using the chain rule, we get $\dfrac{dg}{dy} = g(y)\log t$. Since $g(0) = 1$, we have

$$\frac{\mathrm{d}g}{\mathrm{d}y}(0) = \log t.$$

In words: the limit of $F_{y-1}(t)$ as y tends to zero is $\log t$. Since $F_{-1}(t) = \log t$, this shows that $F_c(t)$ depends continuously on c at $c = -1$!

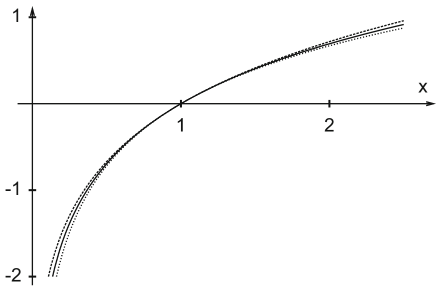

Fig. 6.22 The graphs of $\log x$ and $\frac{x^{c+1}-1}{c+1}$ for $c = -0.9$ and $c = -1.1$

In the next example, we illustrate how to use the fundamental theorem of calculus to construct a function with a particular derivative.

The Logarithm and Exponential Functions Redefined. Suppose we had not worked hard in Chaps. 1 and 2 to define e^x and $\log x$. Let us use the fundamental theorem part (a) to define a function $F(x)$ whose derivative is $\frac{1}{x}$. Let

$$F(x) = \int_1^x \frac{1}{t}\,\mathrm{d}t$$

for any $x > 0$. See Fig. 6.23. Then by the fundamental theorem,

$$F'(x) = \frac{1}{x}.$$

Recall that the derivative of $\log x$ is also $\frac{1}{x}$. Two functions that have the same derivative on an interval differ by a constant, and since both are 0 at $x = 1$, we see that

$$\log x = \int_1^x \frac{1}{t}\,\mathrm{d}t, \qquad x > 0. \tag{6.17}$$

Next we show that if we take Eq. (6.17) as the definition of the logarithm function, we can derive all properties of the logarithm from Eq. (6.17). The basic properties of $\log x$ are:

(a) $\log 1 = 0$,
(b) $\log x$ is an increasing function, and
(c) $\log(ax) = \log a + \log x$.

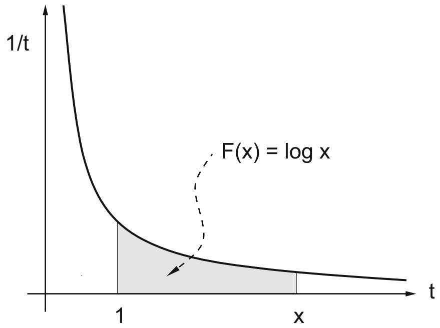

Fig. 6.23 For $x > 1$, $\log x$ can be visualized as the area under the graph of $\frac{1}{t}$

Part (a) follows because the lower limit of integration in Eq. (6.17) is 1. Part (b) follows because the derivative of $\log x$ is positive for x positive. For part (c), take the derivative of $\log(ax)$. Using the chain rule, we get

$$(\log ax)' = \frac{1}{ax}a = \frac{1}{x}.$$

This shows that $\log(ax)$ and $\log x$ have the same derivative on $(0,\infty)$. Therefore, they differ by a constant, $\log(ax) = C + \log x$. Setting $x = 1$, we see that C is $\log a$.

Since $\log x$ is increasing, we can define e^x to be the inverse of $\log x$:

$$\log e^x = x.$$

The basic properties of e^x are:

(a) $e^0 = 1$,
(b) $\left(e^x\right)' = e^x$, and
(c) $e^{a+x} = e^a e^x$.

Part (a) follows from property (a) of the log function. For part (b), note that $\log x$ is differentiable with continuous nonzero derivative. Its inverse, e^x, is then also differentiable. By the chain rule,

$$1 = x' = \left(\log e^x\right)' = \frac{1}{e^x}\left(e^x\right)'.$$

Multiply both sides by e^x to get $(e^x)' = e^x$. For part (c), we first verify that the derivative of $\frac{e^{a+x}}{e^x}$ is zero. By the quotient rule and chain rule,

$$\left(\frac{e^{a+x}}{e^x}\right)' = \frac{e^x e^{a+x} - e^{a+x}e^x}{(e^x)^2} = 0 \qquad \text{for all } x.$$

This means that $\frac{e^{a+x}}{e^x}$ is constant. Taking $x = 0$, we see that the constant is e^a.

Defining e^x as the inverse of $\log x$ is much, much simpler than the definition of e^x we used in Chap. 2: $e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$. Finding $(e^x)'$ as the derivative of the inverse function of $\log x$ is much, much simpler than the discussion in Sect. 3.3a.

Many arguments are much easier once you know calculus. In Problem 6.22, we ask you to use calculus to find a much easier proof that $\left(1+\frac{1}{n}\right)^n$ is an increasing function of n than the one we gave in Sect. 1.4, which used no calculus.

Trigonometric Functions Redefined. In Sect. 2.4, we gave a geometric definition of trigonometric functions and then defined the inverse trigonometric functions from the trigonometric functions. We now show how their inverse functions can be defined independently, using integration. Then the trigonometric functions can be defined as *their* inverses. For $0 < x < 1$, let

$$F(x) = \int_0^x \frac{1}{\sqrt{1-t^2}}\,dt. \tag{6.18}$$

Since $F'(x) = \frac{1}{\sqrt{1-x^2}}$ is positive, $F(x)$ is an increasing function of x for $0 < x < 1$, and therefore F has an inverse. We define this inverse to be $x = \sin t$, defined for $0 < t < p$, where $p = F(1)$ is defined as the limit of $F(x)$ as x tends to 1. All properties of the sine function can be deduced from this definition. In Sect. 7.3, we will see that $F(x)$ does have a limit as x approaches 1. To define the sine function to have the full domain we expect, we have to do a bit more work. But we see that $\sin t$ can be completely described without reference to triangles!

Problems

6.15. Use the fundamental theorem of calculus to calculate the derivatives.

(a) $\frac{d}{dx}\int_0^x t^3\,dt$

(b) $\frac{d}{dx}\int_0^x t^3 e^{-t}\,dt$

(c) $\frac{d}{ds}\int_{-2}^{s^2} x^3 e^{-x}\,dx$

(d) $h'(4)$, if $h(x) = \int_1^x \sqrt{t}\cos\left(\frac{\pi}{t}\right)dt$

6.16. Use the fundamental theorem to calculate the integrals.

(a) $\int_1^2 t^3\,dt$

(b) $\int_0^b (x^3+5)\,dx$

(c) $\displaystyle\int_0^1 \frac{1}{\sqrt{1+t}}\,\mathrm{d}t$

(d) $\displaystyle\int_0^7 \left(\cos t + (1+t)^{1/3}\right)\,\mathrm{d}t$

(e) $\displaystyle\int_a^b (t - \mathrm{e}^t)\,\mathrm{d}t$

6.17. Use the fundamental theorem to calculate the integrals.

(a) $\displaystyle\int_0^{\pi/4} \frac{1}{1+x^2}\,\mathrm{d}x$

(b) $\displaystyle\int_0^1 (x^2+2)^2\,\mathrm{d}x$

(c) $\displaystyle\int_1^4 \left(\frac{2}{\sqrt{x}} - \sqrt{x}\right)\,\mathrm{d}x$

(d) $\displaystyle\int_{-2}^{-1} (2 + 4t^{-2} - 8t^{-3})\,\mathrm{d}t$

(e) $\displaystyle\int_2^6 \left(2s + \frac{1}{s+1}\right)\,\mathrm{d}s$

6.18. Sketch the regions and find the areas.

(a) The region bounded by $y = \sqrt{x}$ and $y = \frac{1}{2}x$.
(b) The region bounded by $y = x^2$ and $y = x$.
(c) The region bounded by $y = \mathrm{e}^x,\ y = -x+1$, and $x = 1$.

6.19. For the function $x(t) = \sin t$ defined as the inverse of F in equation (6.18), show that

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sqrt{1-x^2}.$$

6.20. Suppose f is an even function, i.e., $f(t) = f(-t)$, and set $g(x) = \displaystyle\int_0^x f(t)\,\mathrm{d}t$. Explain why g is odd, i.e., $g(x) = -g(-x)$.

6.21. Suppose g is a differentiable function and

$$F(x) = \int_a^{g(x)} f(t)\,\mathrm{d}t.$$

Explain why $F'(x) = f(g(x))g'(x)$.

6.22. Explain the following items, which prove that $\left(1+\dfrac{1}{n}\right)^n$ is an increasing sequence.

(a) For $x > 0$, $\left(1+\dfrac{1}{x}\right)^x = \mathrm{e}^{x\log\left(1+\frac{1}{x}\right)}$

(b) $\displaystyle\int_1^{1+1/x} \frac{1}{t}\,\mathrm{d}t > \frac{1}{x}$

(c) $\displaystyle\int_1^{1+1/x} \frac{1}{t}\,\mathrm{d}t - \frac{1}{x+1} > 0$

(d) $\displaystyle\frac{\mathrm{d}}{\mathrm{d}x}\left(1+\frac{1}{x}\right)^x > 0$

(e) When n is a positive integer, $\displaystyle\left(1+\frac{1}{n}\right)^n$ is an increasing function of n.

6.23. Use the fundamental theorem to explain the following.

(a) $\displaystyle\frac{1}{4}\left((1+3^3)^4-(1+2^3)^4\right) = \int_2^3 3t^2(1+t^3)^3\,\mathrm{d}t$

(b) $\displaystyle v(t_2)-v(t_1) = \int_{t_1}^{t_2} a(t)\,\mathrm{d}t$ for the acceleration and velocity of a particle.

6.24. Work, or rework, Problems 6.11, 6.12, and 6.13 using the fundamental theorem.

6.4 Applications of the Integral

6.4a Volume

The volume of most regions in three-dimensional space is best described using integrals of functions of more than one variable, a topic in multivariable calculus. But some regions, such as solids of revolution and solids obtained by stacking thin slabs, can be expressed as integrals of functions of a single variable. For example, volumes of revolution can be expressed as quantities that satisfy the two basic properties of additivity and lower and upper bounds.

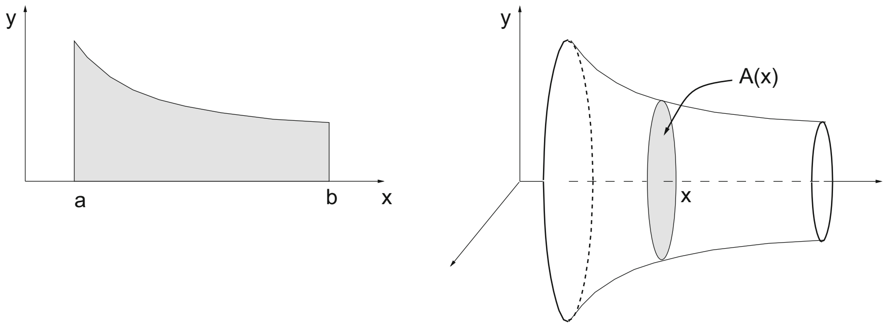

Fig. 6.24 *Left*: A planar region to revolve around the x-axis. *Right*: The solid of revolution. $A(x)$ is the shaded cross-sectional area of the solid at x

Let us write $V(A,[a,b])$ for the volume of a solid of revolution that is located in the region $a \le x \le b$, and where $A(x)$ is the cross-sectional area of the solid at each x. The solid is obtained by revolving a planar region as in Fig. 6.24 around the x-axis, and all cross sections are circular.

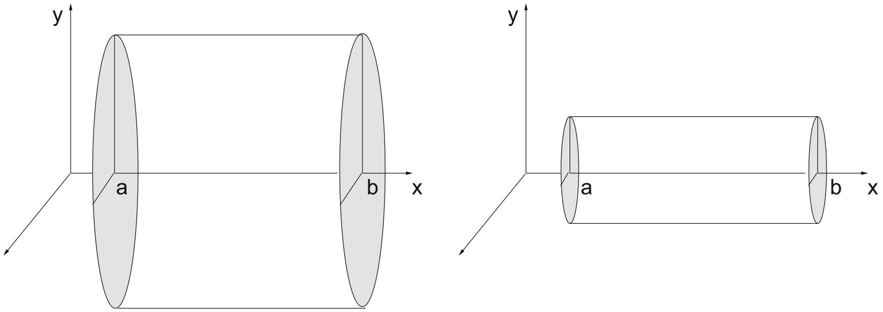

Fig. 6.25 If $m \le A(x) \le M$, the solid in Fig. 6.24 fits between cylinders with volumes $M(b-a)$ and $m(b-a)$

If M and m are the largest and smallest cross-sectional areas,

$$m \le A(x) \le M,$$

then the solid fits between two cylinders having cross-sectional areas m and M, and for the volumes of the three solids, we have (Fig. 6.25)

$$m(b-a) \le V(A,[a,b]) \le M(b-a).$$

Similarly, if we cut the object at c between a and b, we expect that the volumes of the two pieces add to the total:

$$V(A,[a,b]) = V(A,[a,c]) + V(A,[c,b]).$$

These two properties are the additivity and the lower and upper bound properties. So it must be that for a volume of revolution,

$$V(A,[a,b]) = \int_a^b A(x)\,\mathrm{d}x. \tag{6.19}$$

Example 6.13. A ball of radius r is centered at the origin. We imagine slicing it with planes perpendicular to the horizontal axis. In Fig. 6.26, we see that the cross section at x is a circular disk of radius $\sqrt{r^2-x^2}$, so the cross-sectional area is

$$A(x) = \pi(r^2-x^2).$$

Then the volume is

$$\int_{-r}^{r} A(x)\,\mathrm{d}x = \int_{-r}^{r} \pi(r^2-x^2)\,\mathrm{d}x = \pi\left[r^2x-\frac{x^3}{3}\right]_{-r}^{r} = 2\pi\left(r^3-\frac{r^3}{3}\right) = \frac{4}{3}\pi r^3.$$

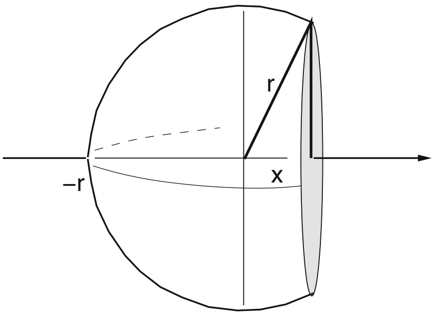

Fig. 6.26 By the Pythagorean theorem, the cross section of a ball is a disk of radius $\sqrt{r^2 - x^2}$, in Example 6.13

Example 6.14. Consider the planar region bounded by the graph of $y = \frac{1}{x}$, the x-axis, and the lines $x = 1$ and $x = 2$. Generate a solid by revolving that region around the x-axis, as in Fig. 6.27. The volume is

$$\int_1^2 A(x)\,\mathrm{d}x,$$

where $A(x)$ is the cross-sectional area of the solid at x. The cross section is a disk of radius $\frac{1}{x}$, so $A(x) = \pi x^{-2}$. The volume is

$$\int_1^2 \pi x^{-2}\,\mathrm{d}x = -\pi x^{-1}\Big|_1^2 = \pi\left(-\frac{1}{2} + 1\right) = \frac{\pi}{2}.$$

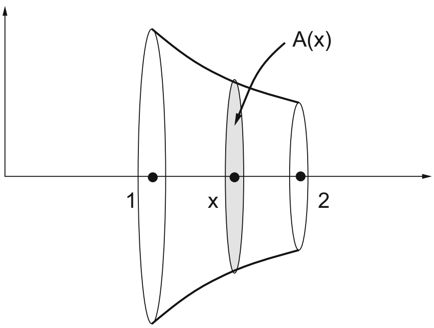

Fig. 6.27 The cross section of the solid in Example 6.14 is a disk of radius $\frac{1}{x}$

6.4b Accumulation

The fundamental theorem of calculus has two important consequences. One is that every continuous function f on an interval arises as the rate of change of some function:

$$f(x) = \frac{d}{dx}\int_a^x f(t)\,dt.$$

The other is that the integral of the rate of change F' equals the change in the function F between a and b:

$$F(b) - F(a) = \int_a^b F'(t)\,dt.$$

In this section, we look at ways to use the integral to answer the following question: How much?

Suppose we know that water is flowing into a pool at a rate $f(t)$ that varies continuously with time. How much water flows into the pool between time a and time b? We subdivide the interval of time into n very small subintervals,

$$a = a_0 < a_1 < \cdots < a_n = b,$$

and for each one, we find the rate at which water is flowing at some time t_i during that interval. The product of the rate $f(t_i)$ and the length of time $(a_i - a_{i-1})$ is a good estimate for the amount of water that entered the pool between times a_{i-1} and a_i. Summing all those estimates, we get the approximate integral $\sum_{i=1}^{n} f(t_i)(a_i - a_{i-1})$. We know that in the limit, such approximate integrals converge to $\int_a^b f(t)\,dt$. So the amount of water that accumulates in the pool between time a and time b is

$$\int_a^b f(t)\,dt.$$

The function $F(t) = \int_a^t f(\tau)\,d\tau$ represents the amount of water that accumulates between time a and time t. Note that if we wanted to know how much water is *in* the pool, we would need to know how much water there was at time a, and then

$$(\text{amount at time } t) = \int_a^t f(\tau)\,d\tau + (\text{amount at time } a).$$

6.4c Arc Length

Here is another example of how the integral is used to answer a "how much" question. Let f have a continuous derivative on $[a,b]$. The *arc length* of the graph of f from a to b is the least upper bound of the sum of lengths of line segments joining points on the graph. Let us see how to compute the arc length using an integral (Fig. 6.28).

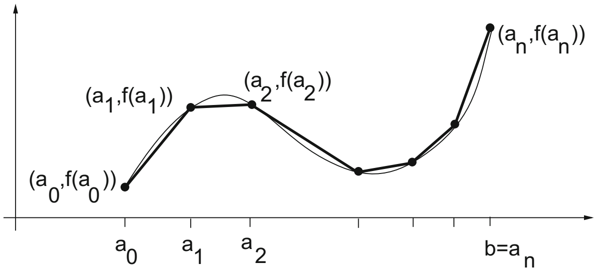

Fig. 6.28 The segments underestimate the arc length

Let

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b$$

be a subdivision of $[a,b]$. By the Pythagorean theorem, the length of the ith segment is

$$\sqrt{(a_i - a_{i-1})^2 + \big(f(a_i) - f(a_{i-1})\big)^2}.$$

By the mean value theorem for derivatives, there is a point t_i between a_{i-1} and a_i such that

$$f(a_i) - f(a_{i-1}) = f'(t_i)(a_i - a_{i-1})$$

and

$$\sqrt{(a_i - a_{i-1})^2 + \big(f(a_i) - f(a_{i-1})\big)^2} = \sqrt{1 + \big(f'(t_i)\big)^2}(a_i - a_{i-1}).$$

The length of the curve is approximately the sum of the lengths of the segments:

$$L \approx \sum_{i=1}^{n} \sqrt{1 + \big(f'(t_i)\big)^2}(a_i - a_{i-1}).$$

Since f' is continuous on $[a,b]$, $\sqrt{1+(f')^2}$ is continuous as well, and the approximate integrals $\sum_{i=1}^{n} \sqrt{1 + \big(f'(t_i)\big)^2}(a_i - a_{i-1})$ approach

$$\int_a^b \sqrt{1 + \big(f'(t)\big)^2}\,\mathrm{d}t,$$

the arc length of the curve from a to b.

Let us see how this formula works on a problem for which we already know the arc length.

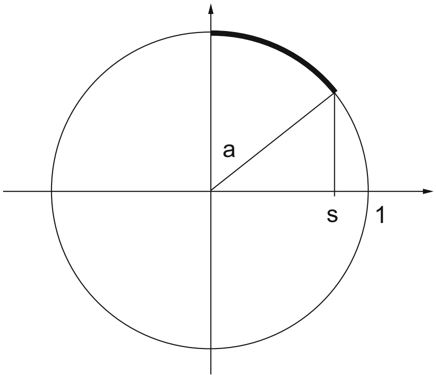

Fig. 6.29 In the first quadrant, the graph of $f(x) = \sqrt{1-x^2}$ is part of the unit circle. See Example 6.15

Example 6.15. According to our definition of the inverse sine function, the length of the heavy arc on the unit circle in Fig. 6.29 ought to be $a = \sin^{-1} s$. Let us check this against the arc-length formula. We have

$$f(x) = \sqrt{1-x^2}, \qquad f'(x) = \frac{-x}{\sqrt{1-x^2}},$$

and

$$\sqrt{1+(f')^2} = \sqrt{1+\frac{x^2}{1-x^2}} = \frac{1}{\sqrt{1-x^2}}.$$

So the arc-length formula gives

$$\int_0^s \frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x = \sin^{-1} s,$$

in agreement with Sect. 3.4c.

Let us compute an arc length that we cannot compute geometrically.

Example 6.16. Find the arc length of the graph of $f(x) = \frac{2}{3}x^{3/2}$ from 0 to 1. We have $f'(x) = x^{1/2}$. Then

$$L = \int_0^1 \sqrt{1+(x^{1/2})^2}\,\mathrm{d}x = \int_0^1 \sqrt{1+x}\,\mathrm{d}x = \frac{2}{3}(1+x)^{3/2}\Big|_0^1 = \frac{2}{3}\left(2^{3/2}-1\right).$$

6.4d Work

The concept of work is readily visualized by the example of hoisting a load with the aid of a rope over a pulley. How much work is required depends on the weight of the load and on the difference between its initial height and the height to which it has to be hoisted. The following facts are suggested in our example by the intuitive notion of work:

(a) The work done is proportional to the *distance* through which the force acts.
(b) The amount of work done is proportional to the weight or force.

Accordingly, we *define* the work, W, done in elevating a load of weight f by the vertical distance h to be

$$W = fh.$$

According to Newton's theory, the dynamical effect of every force is the same. So we take $W = fh$ to be the work done by any force f acting through a distance h in the direction of the force (Fig. 6.30).

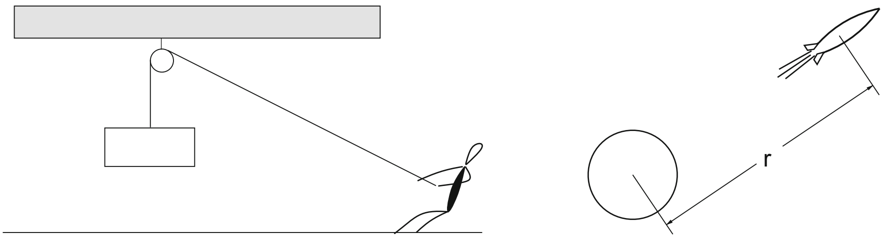

Fig. 6.30 *Left*: Work $W = fh$ to lift weight f through a distance h. *Right*: $W = \int f\,dr$ with a variable force

We show now that this formula for work is meaningful when h is negative, that is, when the displacement is in the opposite direction to the force. For let us lower the load to its original position. The total energy of the load has not changed, so the work done in lowering the load undoes the energy it gained from being raised, and so is a negative quantity.

How much work is done in moving an object through an interval $[a,b]$ against a variable force f, i.e., a force f whose magnitude differs at different points of $[a,b]$ and may even reverse its direction? In this case, f is a function defined on $[a,b]$; let us denote the work done by $W(f,[a,b])$.

What kind of function is W of $[a,b]$? Suppose $[a,b]$ is divided into two disjoint intervals,

$$a < c < b.$$

Since moving the object across $[a,b]$ means moving it first across $[a,c]$, then across $[c,b]$, it follows that the total work is the sum of the work done in accomplishing the separate tasks:

$$W(f,[a,b]) = W(f,[a,c]) + W(f,[c,b])$$

How does W depend on f? Clearly, if at every point of $[a,b]$ the force f stays below some value M, then the work done in pushing against f is less than the work done in pushing against a constant force of magnitude M. Likewise, if the force f is greater than m at every point of $[a,b]$, then pushing against f requires more work than pushing against a constant force of magnitude m. The work done by pushing against a constant force is given by $W = fh$. Thus, if the force f lies between the bounds

$$m \le f(x) \le M \text{ for } x \text{ in } [a,b],$$

then

$$m(b-a) \le W(f,[a,b]) \le M(b-a).$$

We recognize these as the additive and the lower and upper bound properties. These two properties characterize W as the integral

$$W(f,[a,b]) = \int_a^b f(x)\,\mathrm{d}x.$$

Problems

6.25. When a spring is stretched or compressed a distance x, the force required is kx, where k is a constant reflecting the physical properties of the spring. Suppose a spring requires a force of 2000 Newtons to compress it 4 mm. Verify that the spring constant is $k = 500{,}000$ N/m, and find the work done to compress the spring 0.004 m.

6.26. If at time t, oil leaks from a tank at a rate of $R(t)$ gallons per minute, what does $\int_3^5 R(t)\,\mathrm{d}t$ represent?

6.27. Water can be pumped from a tank at the rate of $2t + 10$ L/min. How long does it take to drain 200 L from the tank?

6.28. Find the volume of the solid obtained by revolving the graph of $\frac{2}{3}x^{3/2}$, for $0 \le x \le 1$, around the x-axis.

6.29. Consider the graph of $\frac{1}{x}$ on $[1,2]$.

(a) Set up but do not evaluate an integral for the arc length.
(b) Calculate two approximate integrals I_{approx} for the arc length, using ten subintervals, and taking the t_i at the left, respectively the right, endpoints.

6.30. During the space shuttle program, the shuttle orbiter had a mass of about 10^5 kg.

(a) When a body of mass m is close to the surface of the Earth, the force of gravity is essentially constant,

$$f = mg,$$

with a gravitational constant of $g = 9.8\,\mathrm{m/s}^2$. Use the constant-force assumption $W = mgh$ to calculate the work done against gravity to lift the shuttle mass to a height of 50 m above the launch pad.

(b) When a mass is moved a great distance from the Earth, the force of gravity depends on the distance r from the center of the Earth,

$$f = \frac{GMm}{r^2}.$$

The radius of the Earth is about 6.4×10^6 m. Equating the two expressions for the weight of an object of mass m at the surface of the Earth, find GM.

(c) Calculate the work done against gravity to lift the shuttle orbiter to an altitude of 3.2×10^5 m.

6.31. We have calculated the volume of a solid of revolution as $\int_a^b A(x)\,\mathrm{d}x$, where $A(x)$ is the cross-sectional area at x. We also know that every integral can be approximated with arbitrary accuracy by an approximate integral taken over a small enough subdivision:

$$\int_a^b A(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} A(x_i)\,\mathrm{d}x_i.$$

We observe that each term $A(x_i)\,\mathrm{d}x_i$ is the volume of a thin cylinder with thickness $\mathrm{d}x_i$. Make a sketch to illustrate that the volume of the solid is well approximated by the volume of a stack of thin cylinders.

6.32. The density ρ of seawater changes with the depth. It is approximately $1025\,\mathrm{kg/m}^3$ at the surface, and 1027 at 500 m depth.

(a) Assume that density is a linear function between 0 and 500 m. Find the mass of a column of water that has a uniform cross-sectional area of $1\,\mathrm{m}^2$ and is located between 100 and 500 m deep.

(b) Assume also that $\rho = 1027$ from 500 to 800 m depth. Find the mass of the $1\,\mathrm{m}^2$ column located between 100 and 700 m deep.